IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 4, July-August 2025

Impact Factor: 8.152

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE
SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

IJARETY

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204064

Hybrid Deep-learning Approach for Heart Disease Prediction using CNN and LSTM

Sudeep S, Soumya, Shridhar V

Department of MCA, CMR Institute of Technology, Bengaluru, India

ABSTRACT: Cardiovascular diseases (CVDs) are the main reason for death is due to this particular cause. worldwide, suffering from heart disease alone leads to more than 17.9 million deaths worldwide each year. Early, precise, and accessible diagnostic methods are crucial for reducing mortality rates, and to address these needs, we propose a web-based system for predicting heart disease that integrates applying both ML methods and advanced deep learning approaches to analyse structured clinical data. The system incorporates K-Nearest Neighbours, Random Forest, Logistic Regression, Long Short-Term Memory architectures, and machine support, including Convolutional Neural Networks. Experimental evaluation demonstrated that the Random Forest model achieved 88 % accuracy, 0.91 ROC-AUC, 0.87 F1-Score, while the integration of CNN and LSTM improved recall to 0.88, enhancing early detection capabilities. The proposed platform supports both real-time and batch predictions through a user-friendly interface, making it suitable for deployment in both clinical and resource- limited environments. This work demonstrates the feasibility of deploying AI-driven predictive systems for proactive cardiovascular risk assessment.

KEYWORDS: Cardiovascular Diseases(CVD), Convolutional Neural Networks(CNN), Deep learning approaches, Machine- learning(ML) strategies, and Long Short-Term Memory(LSTM) models, Web-based Application.

I. INTRODUCTION

Cardiovascular diseases (CVDs), or heart disease, are the world's number-one killer, claiming roughly 32 percent of total worldwide deaths, based on data from the World Health Organization[1]. As heart conditions develop due to lifestyle, stress, and aging populations, the healthcare industry is under increasing pressure to create advanced strategies for the prevention and early identification of heart disease. Developments in Artificial Intelligence(AI) and Machine Learning (ML)within healthcare has made unprecedented opportunities for progressing diagnostic precision and advancing in clinical decision-making. Despite advancements in technology, significant challenges remain for identifying heart disease at its initial stages. Conventional diagnostic methods are time-consuming, resource-intensive, and dependent on the availability of skilled clinicians.Most patients living in remote or under-resourced areas do not have access to advanced medical equipment. Further application of current ML models in this setting is seriously uneven due to inconsistency in feature selection and model choice, and information pre-processing. In this way, adaptable, dependable, and user-friendly arrangements are essential to connect specialized capability and real-world usage. To address these limitations, we enhanced our system with deep-learning architectures. CNN was used to identify spatial dependencies in patient health metrics, while LSTM was applied to sequentially process clinical features over time, offering a new dimension to disease pattern recognition[16].

The project investigates the implementation of a machine learning approach for heart disease prediction, building upon traditional diagnostic algorithms such as the one developed by Detrano etal.[2] The arrangement is connected over algorithms for supervised learning techniques such as Support Vector Machines (SVMs) , K-Nearest Neighbors (KNN), Logistic Regression, and Random Forest. The framework runs on organized understanding of well-being data like age, sex, blood, weight ,cholesterol and other clinical parameters benchmarked from datasets such as the Machine-learning Repository[5].The project also includes performance evaluation utilizing standard assessment measurements metrics like F1-score, recall, accuracy, and precision. This study's main contribution is the creation of an end-to-end pipeline from information preprocessing and demonstrate preparation to web deployment to empower real- time forecasting of heart disease. The framework bolsters single and batch expectations with a simple and accessible online interface to medical physicians and analysts. This project could be a strong point of reference within applying machine-learning to support practical clinical decision- making, particularly for low-resource or for-flung clinical environments.

This paper is structured as follows: Part Two offers a comprehensive review of previous work and highlights the issues found in the existing literature. Section 3 explains the method used, Including the process of information preprocessing ,the choice made, and the metrics used to evaluate the results. Section 4 explain how the network application was implemented and configured. The results and their interpretation are covered in Section 5,and a summary of the results and suggestions for further research are provided in Section 6.

II. LITERATURE REVIEW

Heart disease prediction has gained significant attention over the past few years, ML and AI have been increasingly applied in healthcare systems, offering effective methods for early diagnosis, structured treatment strategies, and ultimately reducing global cardiovascular mortality. Recent research has examined multiple Machine-Learning algorithms are applied to improve forecast precision and bolster the dependability of heart disease detection. For instance, Gnaneswari. (2022)[13] highlighted how crucial feature selection is in improving classification outcomes. They demonstrated that selecting the most pertinent clinical features significantly boosts prediction accuracy.Deep-learning has gained prominence in this domain. Garc´ıaOrdas et al. (2024) developed a deep-learning model incorporating feature augmentation achieving impressive results on complex datasets, highlighting the potential of such models to predict severe heart failure outcomes with high accuracy[14]. While powerful, deep-learning models frequently hindered by interpretability challenges, a concern echoed in the systematic review by Fernandes et al. (2020), which noted that explainability remains a barrier to clinical adoption[6].

Multiple researchers have highlighted the weaknesses of existing models and the challenges in real-world deployment. Fernandes et al. (2023)[6] and Chandrasekhar and Pedda Krishna (2023)[15] discussed how inconsistent preprocessing, lack of balanced data, and absence of real-time tools reduce the success of current systems. Gnaneswari (2023)[13] and Zhang et al. (2024)[7] explored data mining and predictive ML models, recommending adoption in healthcare routines for practical use.Our project is motivated by these gaps. While previous studies achieved high model accuracy, most lacked usable web-based platforms for real-time prediction and batch diagnosis. By developing and deploying an integrated Flask- based ML system that supports both CSV input and individual patient assessment, this work addresses both performance and accessibility. The use of Random Forest, proven effective by studies like Patel et al. (2016) [4], Gnaneswari (2022) [13], and Fernandes et al. (2020) [6], forms the core of our prediction engine.

Deep-learning has become a powerful tool for cardiovascular diagnosis. Studies like Garcia-Ordas et al.(2024) demonstrated the role of CNNs in enhancing feature extraction for dependable heart disease prediction[14]. Likewise, LSTM networks are utilized in effectively to model patient data over time, offering improvements in early risk assessment[16]. However, the complexity and lack of transparency in these models remain a challenge for clinical acceptance.To conclude, Although previous studies have shown high predictive accuracy using ML and deep-learning, significant challenges remain, including limited interpretability, lack of deployment in real-time systems, and insufficient evaluation in resource -limited settings. Additionally, few researchers have combined both special feature extraction (via CNN's) and temporal sequence modelling (via LSTM) into a unified, deployable platform.

This research overcomes these restrictions by developing a web-based, real-time heart disease detection system that integrates both CNN and LSTM models for comprehensive spatial-temporal analysis of clinical data, thus improving prediction precision and memory while maintaining usability for healthcare practitioners in diverse settings.

III. METHODOLOGY

This chapter outlines the methodical approach used to create the heart disease prediction framework. The framework allows users to register through an intuitive interface, as displayed in Fig. 1, before accessing the heart prediction platform. Which includes:

- Data collection
- Data preprocessing
- Model building
- Evaluation
- Deployment

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

IJARETY

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204064

Fig1. Register Page of Heart Disease Prediction System

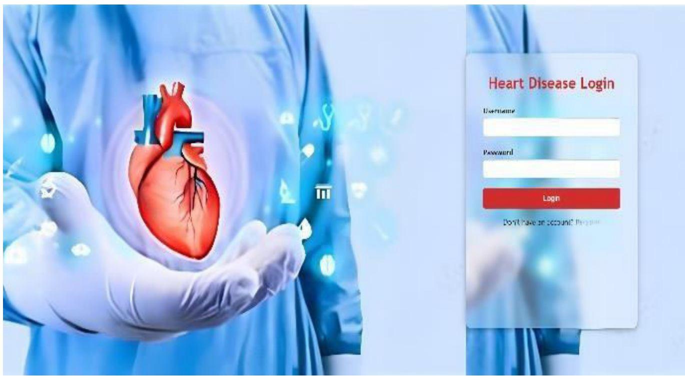

Fig.3. Login Page of the Heart Disease Prediction System

A. Data Description

The system operates on structured clinical information gleaned from public datasets like the Heart Disease dataset or institution-specific health records uploaded in CSV files[5][2]. The dataset contains several features typically employed in cardiac diagnosis, as illustrated in Fig. 2. such as:

Fig2. Dataset Description

Feature	Description
age	Age of the patient
sex	Gender (0 = female, 1 = male)
cp	Chest pain type
trestbps	Resting blood pressure
chol	Serum cholesterol (mg/dl)
fbs	Fasting blood sugar > 120 mg/dl
restecg	Resting electrocardiographic results
thalach	Maximum heart rate achieved
exang	Exercise-induced angina
oldpeak	ST depression induced by exercise
slope, ca, thal	Various ECG and angiographic measures
target	Presence (1) or absence (0) of heart disease

B. Data Ingestion and Storage

A custom-built administrative interface enables healthcare professionals to upload CSV files containing patient data. Uploaded files are parsed and validated before being stored in a relational database (e.g., PostgreSQL or SQLite). This architecture supports scalability and data integrity. Users can log in to the system using a secure authentication interface, as shown in Fig. 3, to upload and manage patient data.

C. Data Preprocessing

To ensure accurate and effective machine-learning model training, a robust data preprocessing pipeline was implemented. This step is critical, as real-world medical datasets often contain inconsistencies, missing values, and mixed data categories that can greatly influence model performance. There was extensive preprocessing involving missing value handling, normalization, and encoding of categorical variables[8],[11].

Multiple machine-learning methods classifiers like logistic regression, decision trees, random forest, and support vector machines, and K-Nearest Neighbours (KNN), underwent training and validation[3],[4],[7],[12],[13]. To asses model effectiveness, metrics like evaluation parameters Measures such as ROC-AUC, F1-score, recall, accuracy, and precision were employed[11]. The System provides a prediction page, where processed input data is submitted for real-time analysis. To pre-process the dataset for model training and evaluation, the following preprocessing steps were applied:

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

IJARETY

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204064

•**Missing Values Handling:** Missing or null values are replaced using mean imputation or removed depending on the attribute's importance.
•**Normalization:** Features such as chol, trestbps, and thalach are normalized with Min-Max or Z-score normalization in order to maintain coherence between scales.
•**Categorical Encoding:** categorical variables such as cp, thal, and slope transformed into numerical through one- hot or label encoding.
•**Train-Test Split:** Here, the data set is separated into testing and training using the 80:20 ratio to evaluate model performance.

D. Model Development
1. Logistic Regression - A statistical model where the binary dependent variable follows a logistic function. It is effective for tasks involving linearly separable data and interpretable feature contributions

2. Decision Tree: A tree-structured model that recursively splits the dataset into branches based on features using measures like Gini impurity or information Gain.It provides rules for classification.

3. Random Forest: A model that aggregates the predictions of multiple decision trees. This increases accuracy and also minimizes overfitting. It is insensitive to noise and works well with diverse datasets.

4. Support Vector Machine (SVM): A classifier that determines the best hyperplane dividing different data points, suitable for both linear and non-linear tasks.

5. K-Nearest Neighbours (KNN): This method labels a data point by checking the majority class of its closes k neighbours, with distance often measured via Euclidean distance.Each algorithm is built on the pre-processed dataset, and its hyperparameters were selected using cross-validation methods like Grid Search or Random Search[11]. Such methods exhaustively search across combinations of model parameters to ascertain the top-performing group according to the validation metrics. The models are examined using cross-validation techniques and evaluation metrics such as F1-score, recall, accuracy, and precision, and ROC-AUC[11]. Hyperparameter adjustment is carried out to improve each model's efficiency, with Random Forest frequently achieving the best balance of performance and interpretability.

User Interface: A responsive web-based frontend is developed where users, typically healthcare professionals, can enter patient information via a form. Once the data is submitted, it is processed by the trained model, and the output is rendered in real-time, indicating whether the patient is susceptible to heart disease.This modular and layered architecture makes certain that the system is not just robust and accurate but also scalable and adaptable to future expansions. In addition to traditional ML algorithms ,deep-learning models like CNN and LSTM were used to capture spatial and temporal characteristics from clinical data[16].

6. Convolutional Neural Networks(CNN): A 1D CNN was designed to identify spatial relationships among input features like cholesterol, resting ECG, and heart rate. The network consisted built with convolutional and pooling layers, ending with dense layers for binary decision-making. CNN helped improve generalization and pattern detection in complex data.

7. Long Short-Term Memory (LSTM): These networks were applied to process sequential clinical inputs, treating each patient's data as a time series. The LSTM's gated structure enabled the model to retain long-term dependencies, which contributed to better recall and early prediction sensitivity.

E. Model Evaluation
To measure effectiveness, the following metrics were applied:
•Accuracy: Fraction of correct outcomes.
•Precision: TP/(TP + FP)
•Recall: TP/(TP + FN)
•F1 Score: Mean of recall and precision.
•ROC-AUC: Area beneath the ROC curve.
These measures give an overall description of model behaviour, particularly under class imbalance scenarios[15].

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204064

F. Deployment Architecture

The end system is comprised of:

•A backend service (Python, Flask) for data uploads and model inference management.
•A trained model saved with pickle or joblib for fast loading.
•A user-facing interface (HTML/CSS/JavaScript) for data input and prediction retrieval.
• An optional API route for integration into hospital management systems or mobile applications.

IV. RESULTS AND DISCUSSION

To analyse the predictive power of the suggested heart disease system, various ML algorithms were developed and assessed using the pre-processed dataset. The most commonly used algorithms' results are illustrated in the following table. Among those tested algorithms, Random Forest provided the best accuracy (88 percent) and ROC-AUC (0.91) [4],[13]. Logistic Regression and SVM were also efficient with good explainability and generalization[9],[10]. The system's output page displays the results of prediction for each patient, as demonstrated in Fig. 4. Deep-learning techniques, with a focus on LSTM, demonstrated superior performance in detecting early heart disease indicators. CNN provided robust feature extraction, making it appropriate for modelling spatial relationships[16]. However, interpretability and computational demands limit their real-world deployment in low-resource settings.

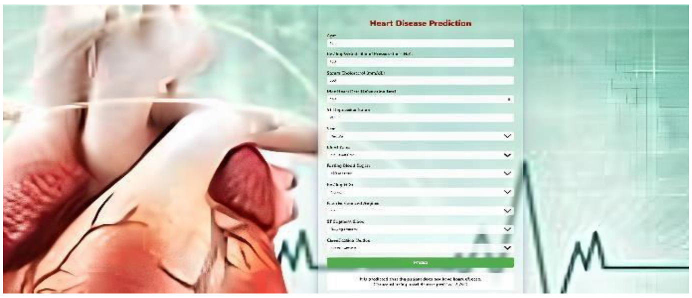

Fig.4. Result Page of the Heart Disease Prediction system

G. Discussion

Random Forest proved most effective because of its capability to handle complex relationships and resist overfitting. The comparative results of various models are presented in TABLE I, highlighting Random Forest's superior performance.

•Logistic Regression provided valuable interpretability, useful in clinical settings.
•Model demonstrated robustness across metrics, suggesting the preprocessing steps and feature selection were appropriate.
•The system's real-time prediction capability through the web interface enables non-technical healthcare personnel to utilize advanced analytics.

Model	Accuracy	Precision	Recall	F1- Score	ROC- AUC
Logistic Regression	82%	0.80	0.78	0.79	0.85
Decision Tree	84%	0.83	0.82	0.82	0.87
Random Forest	**88%**	**0.88**	0.87	0.87	**0.91**
SVM	85%	0.84	0.83	0.83	0.88
KNN	83%	0.81	0.80	0.80	0.86
CNN	86%	0.85	0.84	0.84	0.89
LSTM	87%	0.86	**0.88**	0.87	0.90

TABLE I Comparison of model performances

V. CONCLUSION

Cardiovascular diseases (CVDs) are the principal reason for death globally, with heart disease alone responsible for over 17.9 million deaths annually. This highlights the urgent need for early, precise, and reasonable demonstrative techniques. This project addresses the challenge of forecasting cardiac disease by employing machine-learning (ML) on

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204064

formal persistent health data. Its objective is to progress early discovery and help clinical decision-making. Past studies indicate that concerns such as need for real-time tools, limited accessibility in rural areas, and variable model accuracy. Integrating ML frameworks with real-world, user-friendly interfaces is additionally required. For our arrangement, we created a web-based determining framework utilizing the Carafe system and directed ML models. Our application is able to make real-time and group forecasts through a clear interface and can be used in clinical and personal contexts. Among the models that were tried, the top-ranked model was the Random Forest classifier, beating Logistic Regression, KNN, and SVM.

This study focuses the design and deployment of a methods for predicting heart disease utilizing machine-learning techniques to analyze clinical data using predictive models. The system can serve as an efficient, scalable, and interpretable tool for supporting early diagnosis and risk assessment in healthcare environments. The system's performance was shown to be highly effective, particularly using the Random Forest model, and deployable via a user-friendly web interface for practical use by healthcare professionals. Specifically, the proposed system achieved 88 % accuracy, 0.91 ROC-AUC, and 0.87 F1-Score with Random Forest, while CNN and LSTM models further improved recall to 0.88, enhancing early detection capabilities. This integration of spatial(CNN) and temporal(LSTM) learning within a deployable web-based platform makes this work unique from prior studies, addressing both performance and accessibility challenges in heart disease prediction.The system has been expanded to include CNN and LSTM models, enhancing predictive accuracy and supporting deeper analysis of spatial and temporal data[16]. Future improvements efforts should aim at enhancing the clarity and real-time efficiency of these models.

H. Future Work

•Integration with EHRs (Electronic Health Records) for real-time automated analysis[8].
•Expansion to larger datasets from multiple hospitals to improve generalizability.
•Explainability modules (e.g., SHAP, LIME) to give information on forecasts for clinical trust[10].

REFERENCES

[1] World Health Organization (WHO). "Cardiovascular diseases (CVDs)." https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
[2] Detrano, R., Janosi, A., Steinbrunn, W., et al. "International application of a new probability algorithm for the diagnosis of coronary artery disease." The American Journal of Cardiology, 64(5), 1989.
[3] Gudadhe, M., Wankhade, K., and Dongre, S. "Decision support system for heart disease based on support vector machine and artificial neural network." International Conference on Computer and Communication Technology, 2010.
[4] Patel, J., Tejal, R., and Bhalani, M. "Prediction and analysis of heart disease using ensemble learning techniques." International Journal of Computer Applications, 145(5), 2016.
[5] Dua, D., and Graff, C. "UCI Machine-learning Repository." https://archive.ics.uci.edu/ml/datasets/heart+disease
[6] Fernandes, M., Cardoso, J. S., and Fernandes, P. "Using machine-learning in cardiovascular disease risk prediction: A systematic review." Expert Systems with Applications, 162, 2020.
[7] Zhang, Y., Wang, S., and Wang, L. "A hybrid method for heart disease diagnosis using data mining techniques." Information Technology and Management, 13, 2012.
[8] Chen, H., Hailey, D., Wang, N., and Yu, P. "A review of data quality assessment methods for public health information systems." International Journal of Environmental Research and Public Health, 11(5), 2014.
[9] Lundberg, S. M., and Lee, S.-I. "A unified approach to interpreting model predictions." Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS), 2017.
[10] Ribeiro, M. T., Singh, S., and Guestrin, C. "Why should I trust you?: Explaining the predictions of any classifier." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.
[11] Tan, P.-N., Steinbach, M., and Kumar, V. "Introduction to Data Mining." Pearson Education, 2006.
[12] Likhitha, K. N., et al. "Heart Disease Detection using Machine-learning Technique," Proc. ICESC, pp. 1738–1743, 2021. doi:10.1109/ICESC51422.2021.9532705.
[13] Gnaneswari, G. "Analysis of Heart Disease Parameters and Prediction," Proc. GCAT, pp. 1–6, 2022. doi:10.1109/GCAT55367.2022.9972211.
[14] García-Ordás, M. T., et al. "Deep-learning techniques for heart disease prediction: A comprehensive study." Journal of Biomedical Informatics, vol. 135, 2024.
[15] Chandrasekhar, R., and Peddakrishna, A. "Evaluating Real-Time Heart Disease Risk Models." IEEE Access, 2023.
[16] Jaiswal, A., et al. "Heart Disease Prediction Using LSTM and CNN Models." Computers in Biology and Medicine,2023.

IJARETY

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com